AD 410230

N-63-4-3

410230

MEMORANDUM
RM-3726-PR
JULY 1963

SOME TYPES OF OPTIMAL CONTROL OF STOCHASTIC SYSTEMS

Stuart E. Dreyfus

PREPARED FOR:
UNITED STATES AIR FORCE PROJECT RAND

The RAND Corporation
SANTA MONICA • CALIFORNIA

Errata Sheet for RM-3726-PR (U)

SOME TYPES OF OPTIMAL CONTROL OF STOCHASTIC SYSTEMS

Stuart E. Dreyfus

The statement on page 6 that D-D-D is the minimizing open-loop control sequence is in error. For the example problem given in the Memorandum, the sequence U-U-D is superior. Hence, for this example, the open-loop-optimal feedback scheme discussed on page 8 duplicates the pure feedback scheme.

However, if the arc number associated with going up from A were changed from 5 to 10, the reader can verify that the open-loop-optimal feedback solution still chooses a U decision at A, but optimal pure feedback dictates a D decision with an associated lower expected sum.

Reports Department
The RAND Corporation

MEMORANDUM

RM-3726-PR

JULY 1963

SOME TYPES OF OPTIMAL CONTROL OF STOCHASTIC SYSTEMS

Stuart E. Dreyfus

This research is sponsored by the United States Air Force under Project RAND—contract No. AF 49(638)-700 monitored by the Directorate of Development Planning, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of the United States Air Force.

The RAND Corporation

1700 MAIN ST • SANTA MONICA • CALIFORNIA

PREFACE

Part of the research program of The RAND Corporation consists of basic supporting studies in mathematics, one aspect of which is concerned with optimization processes. This Memorandum is concerned with optimal control of dynamic systems involving random variables. Optimal control rules are developed and evaluated.

Optimization is particularly important in determining rocket trajectories and correcting deviations in flight from the predetermined trajectory.

SUMMARY

The optimal control of stochastic systems is considered. Under various assumptions concerning the information available to the controller, different optimal control rules result. For certain specific problems, the different control schemes are analyzed and compared, and the vast superiority of feedback over open-loop control is demonstrated.

SOME TYPES OF OPTIMAL CONTROL OF STOCHASTIC SYSTEMS

1. INTRODUCTION

A stochastic system (i.e., a dynamic system involving random variables) which evolves according to a rule which also involves variables or parameters under external control, is called a stochastic control system. If these variables or parameters are determined so that the system behaves as well as possible as measured by some well-defined criterion, one has achieved optimal control of the stochastic system.

Under varying assumptions concerning the information available to the controller, different optimal control policies result. In this Memorandum we shall develop and illustrate several different control schemes and compare their behavior. In this way we intend to demonstrate that certain control philosophies that may appear superficially to be equivalent, are really quite different. In the final section we derive asymptotic expressions for the cost of optimal control using several different schemes. This yields a quantitative measure of the vast superiority of feedback over open-loop control for a particular stochastic system.

2. A DETERMINISTIC PROBLEM

Let us begin by considering a trivial three-stage discrete deterministic control problem. Given the directed network shown below,

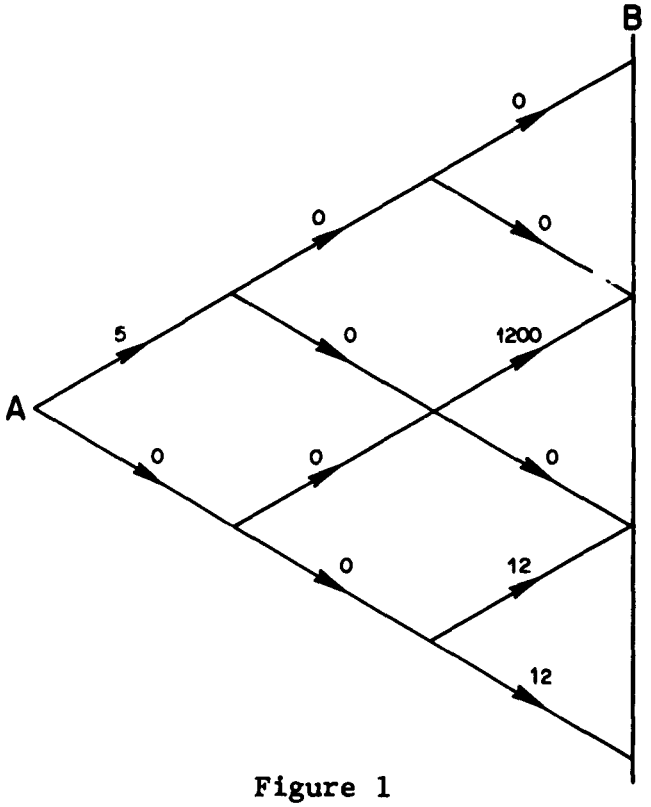

Figure 1

we wish to determine that path from point A to line B which has the minimal sum of the numbers written along the three arcs of the path.

Let us denote a decision to follow the diagonally-up arc from an intersection by U and the diagonally-down arc by D. By examining all eight possible paths from A to B, we discover that the path D-U-D (diagonally down,

then up, then down) has sum-of-arc-numbers zero and is the unique optimal solution. We shall call such a designation of the solution, giving the sequence of control decisions to be followed from specific initial point to termination, the optimal open-loop control

A second way of presenting the solution to this problem is to associate with each node of the figure a decision, either U or D, such that that decision is the initial one of the optimal path from the node to the terminal line. This set of decisions assigned to nodes is most efficiently determined recursively backwards from the terminal line [1]. We initially record the optimal decisions and minimal sum to termination (encircled) at the nodes along the line C in the figure below,

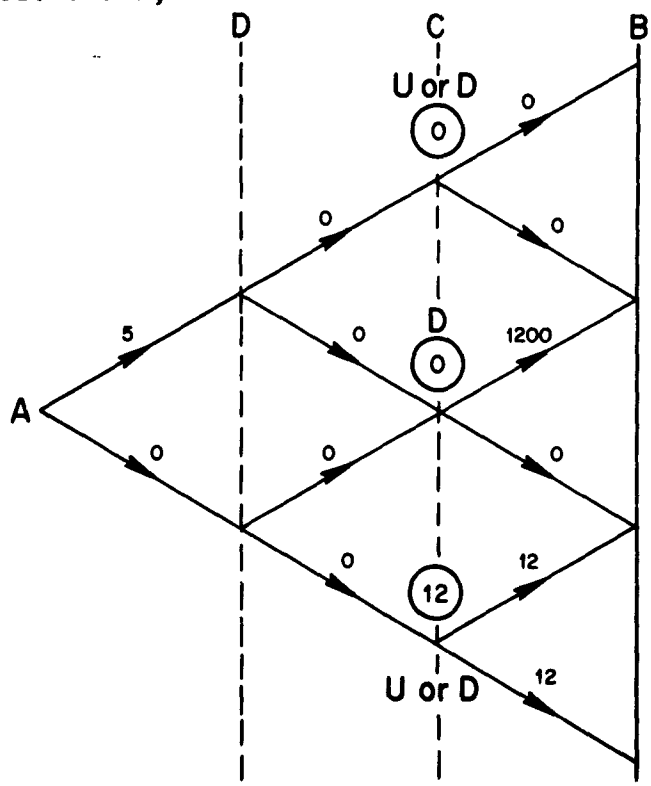

Figure 2

and then use the circled numbers to determine the optimal decisions and sum along D and, finally, from A. The resulting figure is

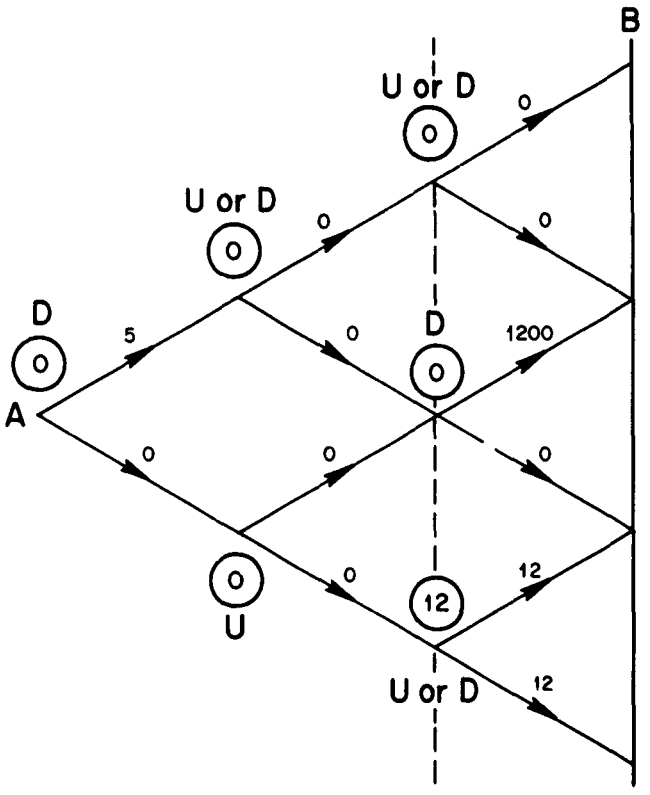

Figure 3

We shall call such a designation of the solution, giving the optimal decision associated with starting at each possible state of the system (i.e., at each node), the feedback optimal control.

The interpretation of Fig. 3 is that the optimal path starting at point A has sum zero and starts diagonally down. The node reached after making the downward move has

a U written by it, indicating a decision to go diagonally up. This leads to a node with a down decision. Hence, D-U-D is the optimal path from A. Note that the feedback representation of the solution also yields the best path starting from other nodes not along the D-U-D path.

The important point is that for a specified initial point such as A, the open-loop and feedback solutions are equivalent for a deterministic process.

3. A STOCHASTIC PROBLEM

Let us now modify the above problem by introducing a stochastic aspect. We shall assume that the decision designated by U results in a probability of 3/4ths of moving diagonally up and 1/4th of moving down. The alternative decision, D, has a 3/4ths chance of a diagonally downward move and a 1/4th chance of an upward transition. We now have a stochastic control problem. We can still exert a controlling influence, but randomness determines the actual transformation of state.

As a criterion for comparing possible control schemes, let us attempt to minimize the expected sum along the path from A to line B.

To determine the best open-loop control policy, we consider all eight possible sequences of decisions and choose the one with minimal expected sum. For example, the decision sequence U-U-U has probability 27/64ths of

actually yielding the path U-U-U with sum 5, 9/64ths probability of yielding the path D-U-U with sum 1200, etc. Multiplying the probabilities by the sums and adding, we get an expected sum E_{UUU} given by

$$E_{UUU} = \frac{27}{64} \cdot 5 + \frac{9}{64} (1200 + 1205 + 5) + \frac{3}{64} (5 + 0 + 12) + \frac{1}{64} \cdot 12 \cong 360.$$

It turns out that the sequence D-D-D has the minimal expected sum of approximately 120.

The best <u>feedback</u> control is computed recursively backwards just as in the deterministic example. Suppose that, for a given node, the expected sums starting at each of the two possible nodes to which one might go have been determined. Then the expected sum from the given node to the termination under decision U is obtained by multiplying the upward arc number plus the remaining expected sum associated with the node at the end of the up-arc by 3/4ths and adding 1/4th times the corresponding downward numbers. Decision D is similarly evaluated reversing the 3/4ths and 1/4th, and the minimal expected sum is chosen. The minimizing decision and expected sum (encircled) are recorded at the node. This computation leads to the figure below:

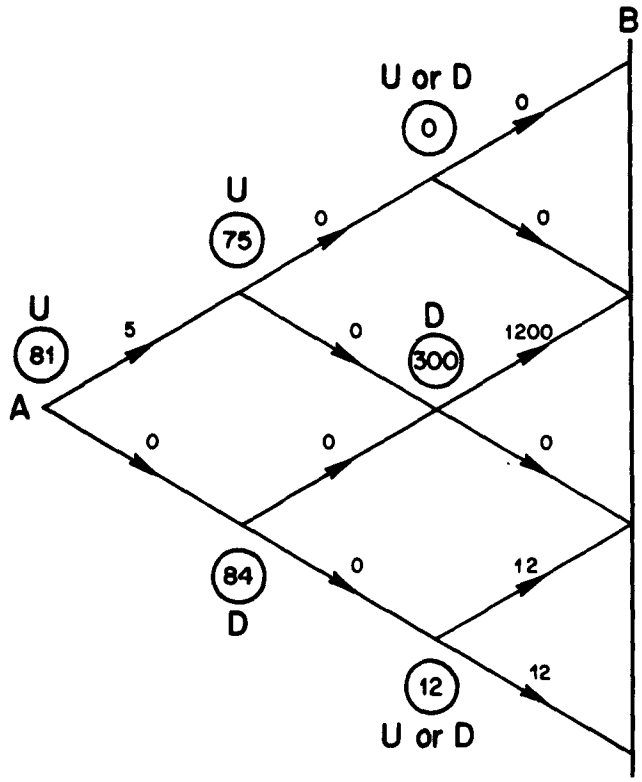

Figure 4

The expected sum using feedback control is 81 and the control policy is the set of letters associated with the nodes in Fig. 4.

At this point we would like to introduce a third control scheme. Let us use the optimal open-loop solution to yield our initial decision. Then, after a transition has occurred, let us observe the result and determine the best open-loop solution for the new two-stage problem. After implementing the initial control decision of this optimal open-loop solution, we again observe the state

and use the optimal control decision for the remaining one-stage problem. This scheme uses the optimal open-loop initial decision at each stage, but incorporates feedback in the observation of the actual state attained. We call this scheme open-loop-optimal feedback control.

This control scheme differs from either of the previous two. The initial optimal open-loop decision agrees with the feedback decision except for starting at node A. There, as has been shown, the optimal open-loop control dictates a downward decision. Therefore, the expected cost of the above scheme is

$$\frac{1}{4} \cdot 80 + \frac{3}{4} \cdot 84 = 83 \ .$$

We can conclude from this example that

1) The pure open-loop scheme incorporating no use of subsequent information about actual transitions yields a large expected sum;
2) The pure feedback scheme where the state is assumed known when the decision is made yields the smallest possible expected sum for a stochastic problem;
3) The open-loop-optimal feedback scheme yields an intermediate expected sum. Although feedback is used, the fact that feedback is to be used is withheld from the computation determining the control decisions, which results in an inferior control scheme.

4. A CONTINUOUS DETERMINISTIC PROBLEM

Let us now consider briefly a standard continuous non-stochastic control problem. Given an initial time t_o and final time T, we wish to use control u(t), $t_o \leq t \leq T$, so as to guide a particle, initially in state x_o, toward the origin $x = 0$. We attach a cost to using control and attempt to minimize the criterion function

$$\int_{t_o}^{T} u^2(t)\,dt + x^2(T) \tag{4.1}$$

where the first term represents the cost of control and the second term measures the deviation from the origin at the terminal time. Motion of the particle is given by the linear differential equation

$$\dot{x}(t) = ax(t) + bu(t) \ . \tag{4.2}$$

This is a linear control problem with quadratic criterion, and has been much analyzed. We consider it briefly here in order to acquaint the reader with the type of problem we shall consider subsequently and with the dynamic programming technique of solution.

The classical necessary conditions for an extremum of the above problem are given in terms of an adjoint variable or Lagrange multiplier λ which satisfies the equation

$$\dot{\lambda} = -a\lambda \tag{4.3}$$

and terminal condition

$$\lambda(T) = 2x(T) \ . \tag{4.4}$$

The optimal control is given by the condition

$$2u + \lambda b = 0 \ . \tag{4.5}$$

Solution of (4.3) with boundary condition (4.4) yields

$$\lambda(t) = 2x(T)e^{a(T-t)} \tag{4.6}$$

and therefore,

$$u(t) = -x(T)be^{a(T-t)} \tag{4.7}$$

so u(t) varies exponentially with time. The unknown terminal value of x, x(T), can be expressed in terms of x(t) by substituting the control rule (4.7) in (4.2) and solving. The resulting expression for x(t) in terms of x(T) can be inverted, and the control at time t is then given in terms of the state at time t by equation (4.7). Performing these steps we get

$$x(t) = x(t_o)e^{a(t-t_o)} + \frac{x(T)b^2}{2a} e^{a(T-t)} - \frac{x(T)b^2}{2a} e^{a(T-2t_o+t)} \tag{4.8}$$

$$x(t_o) = \left(1 - \frac{b^2}{2a} + \frac{b^2}{2a} e^{2a(T-t_o)}\right) e^{-a(T-t_o)} x(T) \tag{4.9}$$

$$x(T) = \frac{e^{a(T-t)} x(t)}{1 - \frac{b^2}{2a} + \frac{b^2}{2a} e^{2a(T-t)}} \tag{4.10}$$

$$u(t) = - \frac{be^{2a(T-t)} x(t)}{1 - \frac{b^2}{2a} + \frac{b^2}{2a} e^{2a(T-t)}} . \tag{4.11}$$

This is the feedback solution for control as a function of state. The optimal control is exponential in time, or, for a given time, it is a linear function of the state.

The dynamic programming solution of this problem proceeds as follows: Define an auxiliary function $f(x,t)$ as the minimal obtainable value of the criterion function (4.1) if we start the problem in state x at time t, $t_o \leq t \leq T$. By the principle of optimality linking the initial decision with the remaining optimal decisions, we have

$$f(x,t) = \min_{u(t)} \left[u^2(t)dt + f(x+(ax + bu)dt,\ t + dt)\right]. \tag{4.12}$$

Expanding (4.12) in Taylor series, dividing by dt and letting dt approach 0, we get

$$0 = \min_{u} \left[u^2 + \frac{\partial f}{\partial x}(ax + bu) + \frac{\partial f}{\partial t}\right]. \tag{4.13}$$

Differentiating with respect to u to minimize gives

$$2u + b\frac{\partial f}{\partial x} = 0 \tag{4.14}$$

and substituting u determined by (4.14) in (4.13), we obtain the non-linear partial differential equation

$$0 = -\frac{b^2\left(\frac{\partial f}{\partial x}\right)^2}{4} + ax\frac{\partial f}{\partial x} + \frac{\partial f}{\partial t}. \tag{4.15}$$

Assuming $f(x,t)$ has the separable form $g(t)x^2$ and substituting in (4.15), we find that $g(t)$ satisfies the Riccati ordinary differential equation

$$-b^2g^2(t) + 2ag(t) + g'(t) = 0 \tag{4.16}$$

with

$$g(T) = 1. \tag{4.17}$$

Solution of this equation yields

$$g(t) = + \frac{e^{2a(T-t)}}{1 - \frac{b^2}{2a} + \frac{b^2}{2a} e^{2a(T-t)}} \tag{4.18}$$

whence

$$f(x,t) = \frac{e^{2a(T-t)} x^2}{1 - \frac{b^2}{2a} + \frac{b^2}{2a} e^{2a(T-t)}} . \tag{4.19}$$

Substitution in (4.14) yields the control scheme

$$u(t) = - \frac{be^{2a(T-t)}}{1 - \frac{b^2}{2a} + \frac{b^2}{2a} e^{2a(T-t)}} \; x(t) \tag{4.20}$$

which agrees with (4.11). Again, as in Sec. 2, we see that for a deterministic problem the open-loop and feedback solutions are equivalent.

5. A CONTINUOUS STOCHASTIC PROBLEM [2-5]

To construct a stochastic control problem, we attach a random variable to the equation defining the evolution of x. We write the discrete rule

$$x(t+\Delta t) = x(t) + [ax(t) + bu(t)]\,\Delta t + \xi(t) \tag{5.1}$$

where $\xi(t)$ is a stochastic process with, for all t,

$$(1)\quad E\,(\xi(t)) = 0 \tag{5.2}$$

$$(2)\quad E\,(\xi^2(t)) = \sigma^2\Delta t \tag{5.3}$$

$$(3)\quad E\,(\xi^n(t)) = o(\Delta t),\ n > 2 \tag{5.4}$$

$$(4)\quad \xi(t_1), \ldots, \xi(t_n) \text{ are independent for any finite collection of distinct times } t_1, \ldots, t_n \tag{5.5}$$

where E is the expected value operator, σ^2 is a constant, and $x = o(\Delta t)$ means the limit as $\Delta t \to 0$ of $\frac{x}{\Delta t}$ is zero. The limiting process as $\Delta t \to 0$ is the continuous control problem we shall consider. Our criterion function to be minimized is

$$E\left[\int_{t_o}^{T} u^2(t)dt + x^2(T)\right], \tag{5.6}$$

the expected cost of control plus terminal deviation.

The optimal <u>open-loop</u> control is deduced by considering all possible functions $u(t)$, $t_o \le t \le T$, and choosing the one that minimizes the criterion (5.6). The cost of control integral is deterministic. Furthermore, if $x(T)$ is viewed, at the initial time t_o, as a random variable dependent upon $u(t)$, one notes that the variance $\sigma^2_{x(T)}$

of this random variable is independent of u(t). Since the expected value of the square of a random variable is its mean squared plus its variance, we have

$$E(x^2(T)) = \left[E(x(T))\right]^2 + \sigma^2_{x(T)} \tag{5.7}$$

so we wish to choose that u(t) which minimizes

$$\int_{t_o}^{T} u^2 dt + \left[E(x(T))\right]^2 . \tag{5.8}$$

Due to the linearity of the equation of evolution (5.1), the expected value of x(T) is the value of x(T) that results from integrating (5.1) with forcing function u(t) and with the stochastic process $\xi(t)$ replaced by its mean value at each time, zero. Hence, our problem reduces, for the special assumptions of linear equations and quadratic criterion, to precisely the deterministic problem that we solved in the previous section.

This observation leads to a fourth control scheme, called certainty equivalent control [6]. This scheme replaces the random variables in the stochastic problem by their expected values and solves the resulting deterministic control problem. Certainty equivalent control is seen to be equivalent to optimal open-loop control in the above example.

To obtain the open-loop-optimal feedback control for

the above problem, we express the control as a function of state, as was done in equation (4.11), and use that control having observed the state transition. The actual realization of the control function then depends upon the realization of the stochastic process; one expects this scheme to perform better than the pure open-loop solution.

The pure feedback control law can be derived by dynamic programming. One defines $f(x,t)$ as the minimal value of (5.6), and writes

$$f(x,t) = \min_u \; E_\xi \left[u^2 \, \Delta t + f(x + (ax + bu) \, \Delta t + \xi, \; t + \Delta t) \right] \quad . \tag{5.9}$$

Hence, expanding in series and taking expectations using (5.2) through (5.5),

$$0 = \min_u \left[u^2 + \frac{\partial f}{\partial x} (ax + bu) + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2} + \frac{\partial f}{\partial t} \right] \quad . \tag{5.10}$$

Therefore,

$$u = - \frac{b \frac{\partial f}{\partial x}}{2} \tag{5.11}$$

and we must solve the equation

$$0 = - \frac{b^2 \left| \frac{\partial f}{\partial x} \right|^2}{4} + ax \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2} + \frac{\partial f}{\partial t} \quad . \tag{5.12}$$

Letting

$$f(x,t) = g(t)x^2 + h(t)$$
$$g(T) = 1 \qquad (5.13)$$
$$h(T) = 0$$

we find that g(t) satisfies the same equation, (4.16), as in the deterministic case. Since the optimal control only involves g(t), we have the same control rule as in Sec. 4, but not the same expected cost, due to the h(t) term reflecting the cost of the randomness. Hence, the optimal feedback control duplicates the open-loop-optimal feedback scheme.

These equivalences of various control schemes are unusual and are the result of our many assumptions of linearity and quadraticity. In the next section we shall modify the problem slightly and demonstrate the dissimilarity of the four different control philosophies we have distinguished.

6. ANOTHER CONTINUOUS STOCHASTIC PROBLEM

We now modify the above problem slightly. We assume that the variance of the random variable ξ in equation (5.1) depends upon the control decision, with no randomness in the evolution of x if no control is exerted. This assumption reflects reality in many applications. We replace (5.3) by the equation

$$E(\xi^2(t)) = u^2\sigma^2\Delta t \tag{6.1}$$

where σ^2 is a constant. We neglect the cost of control integral in the objective function (5.6), since the cost of control is now reflected in the uncertainty attendant upon the use of control. Our criterion function is now merely

$$E\left[x^2(T)\right] \quad . \tag{6.2}$$

For simplicity, we take a = 0 in the equation of evolution (5.1), and use the continuous limit of

$$x(t + \Delta t) = x(t) + [bu(t)]\,\Delta t + \xi(t) \quad . \tag{6.3}$$

We first consider optimal <u>open-loop</u> control. The variance of the random variable x(T) as viewed at time t_o is

$$\int_{t_o}^{T} u^2(t)\,\sigma^2\,dt \tag{6.4}$$

and the criterion function equals

$$\left[E(x(T))\right]^2 + \int_{t_o}^{T} u^2\sigma^2 dt \quad . \tag{6.5}$$

By the same reasoning as above, the expected value of

$x(T)$ is the value yielded by replacing the random variable ξ at each time t by its mean, zero. We therefore have the same problem as in Sec. 4 and Sec. 5, except for a factor σ^2 in the criterion function and no ax term in the equation of motion. The adjoint variable $\lambda(t)$ is, in this case, a constant with terminal value $2E(x(T))$. The optimal control is given by

$$u(t) = - \frac{E(x(T))b}{\sigma^2} \tag{6.6}$$

and is a constant function of time. Expressed in terms of state, we have

$$u(t) = - \frac{x(t)}{b(T-t + \frac{\sigma^2}{b^2})} \tag{6.7}$$

which, as before, is linear in the state at a given time. Using open-loop control, the expected terminal value of x, if we start at time t_o in state $x(t_o)$, is

$$E[x(T)] = \frac{\sigma^2 \, x(t_o)}{b^2(T-t_o + \frac{\sigma^2}{b^2})} \tag{6.8}$$

and the variance of the random variable $x(T)$ is given by

$$\sigma^2_{x(T)} = \frac{\sigma^2 \, x^2(t_o) \, (T-t_o)}{b^2 \, (T-t_o + \frac{\sigma^2}{b^2})^2} . \tag{6.9}$$

Hence, the value of the criterion function is given by

$$E\left[x^2(T)\right] = \left[E(x(T))\right]^2 + \sigma^2_{x(T)} = \frac{\sigma^2\, x^2(t_o)}{b^2(T-t_o+\frac{\sigma^2}{b^2})} \qquad (6.10)$$

We next analyze the open-loop-optimal feedback control scheme. This involves using the rule (6.7) for control as a function of state. The equation of motion becomes

$$x(t+\Delta t) = x(t) - \frac{x(t)}{(T-t+\frac{\sigma^2}{b^2})}\Delta t + \xi(t) \; . \qquad (6.11)$$

If we define $f(x,t)$ as the expected value of $x^2(T)$ using the above rule, we have

$$f(x,t) = \underset{\xi}{E}\left[f(x - \frac{x\,\Delta t}{T-t+\frac{\sigma^2}{b^2}} + \xi \; , \; t+\Delta t)\right] \qquad (6.12)$$

which, after series expansion, letting $\Delta t \to 0$, and taking the expectation, gives

$$0 = -\frac{x}{T-t+\frac{\sigma^2}{b^2}}\frac{\partial f}{\partial x} + \frac{x^2\,\sigma^2}{2b^2(T-t+\frac{\sigma^2}{b^2})^2}\frac{\partial^2 f}{\partial x^2} + \frac{\partial f}{\partial t} \; . \qquad (6.13)$$

Letting $f(x,t)$ have the form

$$f(x,t) = g(t)x^2 \qquad (6.14)$$
$$g(T) = 1$$

we obtain the linear homogeneous equation for $g(t)$

$$g'(t) + \frac{1}{T - t + \frac{\sigma^2}{b^2}} \left[\frac{\sigma^2}{b^2(T - t + \frac{\sigma^2}{b^2})} - 2 \right] g(t) = 0 \qquad (6.15)$$

so that

$$f(x,t) = x^2 \exp\left\{ \int_t^T \frac{1}{T - \tau + \frac{\sigma^2}{b^2}} \left[\frac{\sigma^2}{b^2(T - \tau + \frac{\sigma^2}{b^2})} - 2 \right] d\tau \right\} \qquad (6.16)$$

$$= x^2 \exp\left\{ 1 - \frac{\sigma^2}{b^2(T - t + \frac{\sigma^2}{b^2})} + 2 \log \frac{\sigma^2}{b^2} - 2 \log \left(T - t + \frac{\sigma^2}{b^2}\right) \right\}. \qquad (6.17)$$

To evaluate the expected terminal x value, given that we start in state $x(t_o)$ at time t_o, we can solve equation (6.13) with solution of the form

$$f(x,t) = g(t)x \qquad (6.18)$$
$$g(T) = 1$$

obtaining

$$E\left[x(T)\right] = \frac{\sigma^2\, x(t_o)}{b^2(T - t_o + \frac{\sigma^2}{b^2})} \,. \tag{6.19}$$

This result is the same as the pure open-loop result (6.8), which is explained by the linearity of the process.

Analysis of the <u>feedback</u> scheme begins with the definition of $f(x,t)$ as the value of the criterion if we start in state x at time t, $t_o \le t \le T$, and use an optimal policy. By the principle of optimality, we have

$$f(x,t) = \min_u \underset{\xi}{E}\left[f(x + (bu)\,\Delta t + \xi\,,\ t + \Delta t)\right] \tag{6.20}$$

which yields

$$0 = \min_u \left[bu\,\frac{\partial f}{\partial x} + \frac{u^2\sigma^2}{2}\,\frac{\partial^2 f}{\partial x^2} + \frac{\partial f}{\partial t}\right] \,. \tag{6.21}$$

Hence, setting the derivative with respect to u equal to zero to minimize,

$$u = -\frac{b\,\frac{\partial f}{\partial x}}{\sigma^2\,\frac{\partial^2 f}{\partial x^2}} \tag{6.22}$$

and, substituting (6.22) in (6.21),

$$0 = -\frac{b^2}{2\sigma^2}\,\frac{\left(\frac{\partial f}{\partial x}\right)^2}{\frac{\partial^2 f}{\partial x^2}} + \frac{\partial f}{\partial t} \quad . \tag{6.23}$$

Setting

$$f(x,t) = g(t)x^2 \tag{6.24}$$
$$g(T) = 1$$

we get

$$0 = -\frac{b^2}{\sigma^2}\,g(t) + g'(t) \quad . \tag{6.25}$$

Solving for $g(t)$,

$$f(x,t) = e^{-\frac{b^2}{\sigma^2}(T-t)}\,x^2 \tag{6.26}$$

$$u = -\frac{bx}{\sigma^2} \quad . \tag{6.27}$$

If we now define $h(x,t)$ to be the expected terminal x value starting in state x at time t and using control (6.27), we can characterize $h(x,t)$ by

$$h(x,t) = \underset{\xi}{E}\left[h\left(x - \frac{b^2x}{\sigma^2}\,\Delta t + \xi,\ t + \Delta t\right)\right] \tag{6.28}$$

where the boundary condition is now

$$h(x,T) = x \ . \tag{6.29}$$

Letting

$$h(x,t) = g(t)x \tag{6.30}$$
$$g(T) = 1$$

we find

$$h(x,t) = e^{-\frac{b^2}{\sigma^2}(T-t)} x \ . \tag{6.31}$$

The final control philosophy we have mentioned above is <u>certainty equivalent</u> control, the optimal control for the deterministic system that results from replacing all random variables in the stochastic problem by their expected values. This yields the problem: Choose u(t) so that x(T) given by

$$\dot{x}(t) = bu(t) \tag{6.32}$$
$$x(t_o) = x_o$$

minimizes the expression

$$x^2(T) \ . \tag{6.33}$$

A little reflection shows that $x(T)$ can be made zero by any of an infinite class of controls, and the problem is therefore not meaningful.

We are now in a position to recapitulate our results. Foremost is the conclusion that the four different control schemes give four different optimal control rules. For open-loop control we have a rule given as a function of time and, naturally, dependent upon t_o, $x(t_o)$, and T. This rule, which never depends upon the realization of the stochastic process and which, in our particular example, is a constant function of time, is (by equations (6.6) and (6.8))

$$u(t) = - \frac{x(t_o)}{b(T-t_o + \frac{\sigma^2}{b^2})} . \tag{6.34}$$

The open-loop-optimal feedback control law is expressed as a function of current state and time and depends upon the realization of the stochastic process. It does not depend explicitly on the initial state or time. This law is (equation (6.7))

$$u(t) = - \frac{x(t)}{b(T-t + \frac{\sigma^2}{b^2})} . \tag{6.35}$$

Note that this law is the same as (6.34) initially (for state $x(t_o)$ at time t_o) and that it duplicates (6.34) if

and only if the stochastic process takes on its mean value, zero. The feedback control law depends on the current time and state, just as does the above scheme. However, due to the fact, stressed earlier, that the optimization mathematics is aware of the feedback nature of the control, we get a law different from (6.35); namely (equation 6.27)

$$u(t) = -\frac{bx(t)}{\sigma^2} \tag{6.36}$$

which, in this particular case, does not happen to depend explicitly on the current time. The certainty equivalence concept, as noted earlier, is inappropriate here and yields no unique control law.

If we examine the asymptotic behavior of the criterion function for a long process ($T \to \infty$) starting at time zero in state x_o, we see that the expected value of $x^2(T)$ approaches zero in all cases. This is because for a long process very little control is exerted at any particular time, hence there is little randomness and we can steer assuredly toward the origin. The nature of the approach to zero as a function of the length of the process, T, is significant. For open-loop control the approach is inverse-linear, with (equation 6.10)

$$E\left[x^2(T)\right] \sim \frac{\sigma^2 x_o^2}{b^2} T^{-1} . \tag{6.37}$$

For open-loop-optimal feedback control we have inverse-square convergence, with (equation 6.17)

$$E\left[x^2(T)\right] \sim \frac{\sigma^4 e x_o^2}{b^4} T^{-2} \quad . \tag{6.38}$$

Finally, the feedback control scheme yields negative-exponential convergence (equation 6.26)

$$E\left[x^2(T)\right] \sim e^{-\frac{b^2}{\sigma^2} T} x_o^2 \quad . \tag{6.39}$$

Both the open-loop and open-loop-optimal feedback schemes can be expected to reach the same terminal x value (equations 6.8 and 6.19), but due to its feedback nature, the latter scheme has less variance associated with it. The pure feedback control has an expected terminal value much closer to the origin (equation 6.31) since one can aim closer with the assurance that deviations resulting from the randomness caused by the greater control will be corrected later. Examining the control rules themselves for a fixed initial point, one finds that the pure feedback scheme calls for greater control. This can be explained by the fact that the feedback scheme can afford to aim closer to the origin in the assurance that overshooting due to randomness can be caught and corrected. While the open-loop-optimal feedback scheme will also catch and correct overshoot, the computation of the control rule is not

cognizant of this fact and is, therefore, more conservative. Pure open-loop control, of course, will not compensate.

7. CONCLUSION

We see than that for any but the simplest stochastic problems, the various control philosophies that are equivalent for deterministic problems are quite dissimilar. Further, we have obtained some quantitative idea of the relative behavior and performance of several different optimal control schemes.

REFERENCES

1. Dreyfus, S. E., "Dynamic Programming," Chap. 5, Progress in Operations Research, Vol. 1, R. L. Ackoff, ed., John Wiley and Sons, New York, 1961.

2. Bellman, R. E., Adaptive Control Processes: A Guided Tour, Princeton University Press, Princeton, New Jersey, 1961.

3. Florentin, J. J., "Optimal Control of Continuous Time, Markov, Stochastic Systems," J. Electronics & Control, Vol. X, No. 6, June 1961, pp. 473-488.

4. Kushner, H. J., "Optimal Stochastic Control," Correspondence, IRE Transactions on Automatic Control, October 1962, pp. 120-122.

5. Fleming, W. H., "Some Markovian Optimization Problems," J. Math. & Mech., Vol. 12, No. 1, January 1963, pp. 131-140.

6. Theil, H., "A Note on Certainty Equivalence in Dynamic Planning," Econometrica, Vol. 25, No. 2, April 1957, pp. 346-349.